

In Silico Analysis of Crustacean Hyperglycemic Hormone Family

S.H. Chen,¹* C.Y. Lin,²* C.M. Kuo³

¹Institute of Zoology, Academia Sinica, Taipei, Taiwan

²Division of Biostatistics and Bioinformatics, National Health Research Institutes, Taipei, Taiwan

³Marine Research Station, Institute of Zoology, Academia Sinica, No. 23-10, Dahy-Uen Rd., Jiaw-Shi, Ilan, Taiwan

Received: 5 February 2004 / Accepted: 26 June 2004 / Online publication: 8 June 2005

Abstract

Through multiple sequence alignment and phylogenetic analysis, the subgrouping of the crustacean hyperglycemic hormone (CHH) family was updated using the most complete, nonredundant sequence data set. All sequences from insects were clustered into a distinct subbranch with characters closer to CHH subfamily I. Several sequences that are controversial in their nomenclature and classification are discussed. The motif configuration of CHHs differs from that of molt-inhibiting hormone or gonad-inhibiting hormone in both N and C termini. These two motifs approach each other in tertiary structure models, and the motif preference reveals the critical roles of these regions in functional specificity. Two types of exon organizations of the CHH family genes were observed. Four-exon *Chh* genes were found in a wide range of pan-crustacean (crustacean and hexapod) taxa, except for the penaeid species, from which the 3-exon *Chh* genes were reported. Meanwhile, the 3-exon structure was found in the *Mih* gene and *Moih* genes from one brachyuran species. Combining gene scan skill and exon splicing rules found in this study, we define three more novel sequences from two insect genomes. The pattern of the exon-exon junction within the mature peptide segment is preserved in all CHH family members.

Key words: CHH family — invertebrate neurohormone — bioinformatics — phylogenetic analysis — motif analysis — tertiary structure — gene structure

Introduction

Studies on the crustacean X organ-sinus gland complex (XOSG) have made great progress in increasing understanding of the physiologic roles and molecular nature of the neuropeptides secreted from it in recent years. The neuropeptides, including crustacean hyperglycemic hormone (CHH), molt-inhibiting hormone (MIH), gonad-inhibiting hormone (GIH), or vitellogenesis-inhibiting hormone (VIH), and mandibular-organ-inhibiting hormone (MOIH), were reported in various decapod species. These neuropeptides are considered to be long to the CHH family because of their sequence similarity (de Kleijn and Van Herp, 1995; Soye, 1997; Keller et al., 1999; Huberman, 2000; Bocking et al., 2001). Besides these, ion transport peptide (ITP) and ITP-like (ITP-L), identified in insects, were also found to be similar to the CHH family peptides.

CHH, primarily involved in carbohydrate metabolism, is released in response to a variety of internal and external signals to meet the energy needs of physiologic compensation. MIH inhibits molting hormone (ecdysteroid) secretion from Y organs. Therefore, a molting cycle is initiated when the circulating MIH concentration diminishes, or its secretion stops. GIH is identified by its role in inhibiting vitellogenesis in females. MOIH represses the synthesis of methyl farnesoate (MF) by the mandibular organ, which is involved in the female ovary growth and morphogenesis of the male propodus (Laufer et al., 1998, 2002). ITP simulates the transport of ions across the ileal epithelium in locusts, while the function of ITP-L remains to be clarified (Audsley et al., 1992; Macins et al., 1999). Two CHH isoforms, CHH A and CHH B, have been identified from American lobster *Homarus americanus* for their hyperglycemic function (Chang et al., 1990): CHH B stimulated oocyte growth (Tensen et al., 1989), while CHH A showed the potential to regulate molting (Chang et al., 1990). Khayat

*The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint first authors.

Correspondence to: C.M. Kuo; E-mail: cmkuo@gate.sinica.edu.tw

et al. (1998) demonstrated neuropeptides from *Marsupenaeus japonicus* with dual roles of CHH and GIH. Seven CHH family neuropeptides with hyperglycemic potency were purified from the sinus gland extract by high-performance liquid chromatography (HPLC) and 6 of them were able to inhibit protein and messenger RNA synthesis in ovarian fragments of *Penaeus semisulcatus* in vitro (Khayat et al., 1998). The involvement of CHH in branchial ionic transport in the crab *Pachygrapsus marmoratus* was reported elsewhere (Spanings-Pierrot et al., 2000).

CHH family neuropeptides range from 8 to 11 kDa and are 72 to 87 amino acid residues in length. Studies on the primary structure have revealed 6 conserved cysteine residues forming 3 intramolecular disulfide bonds; the cysteine pairs in disulfide bonds are arranged in a (1,5), (2,4), (3,6) fashion (Keller, 1992; Yasuda et al., 1994; de Kleijn et al., 1998). Common sequence characters among the CHH family members have been described in several ways, such as the regular expression of the crustacean CHH_MIH_GIH neurohormone family signature in the PROSITE entry PS01250, [LIVM]-x(3)-C-[KR]-x-[DENGHRH]-C-[FY]-x-[STN]-x(2)-F-x(2)-C; the position-specific scoring matrix in the Pfam entry, PF01147 (Crust_neurohorm); and the conserved sequence blocks in the PRINTS entry, PR00550 (HYPRGLYCEMIC). These methods are summarized in INTERPRO entry IPR001166, the crustacean neurohormone CHH/MIH/GIH. The CHH family is further separated into CHH subfamily I (or the CHH subfamily) and CHH subfamily II (or the MIH/GIH subfamily), according to sequence homology and functional divergence (Keller, 1992; de Kleijn and Van Herp, 1995; Lacombe et al., 1999; Bocking et al., 2001). These two subfamilies are also described in INTERPRO entry IPR000346 (Hyperglycemic hormone, type 1) and IPR001262 (Hyperglycemic hormone, type 2). Multiple sequence alignment and motif searching on 32 CHH-related sequences were applied to characterize molecules independently of their biological functions (Lacombe et al., 1999). These sequences were segregated into two subgroups with 5 short motifs each. The motifs A1 and A1' were associated with functional specificity, while motifs A5 and A5' were suggested to determine species specificity. However, the functional domains they defined were determined by some sequences of fewer taxa, which may mean the conclusion is less representative and not applicable to other sequences.

Recently, an increasing number of CHH-related sequences have been deciphered, providing valuable information. In this study, bioinformatic approaches

were applied to the updated and nonredundant CHH-related sequences from public databases and the published reports. Multiple sequence alignment and phylogenetic analyses were used to reexamine the subgrouping of these sequences. Correlation of motifs to biological activities was a major focus. The results may serve as a basis of experimental designs to study the function of these molecules. The gene organizations of the CHH family sequence were also analyzed and discussed. Through use of data-mining skills and knowledge of the *Chh*-related gene structure, we discovered three novel CHH-related sequences from two insect genomes.

Materials and Methods

Collection of CHH Family Sequences. Protein sequences used in this study were extracted from a nonredundant protein database (GenBank CDS translations + PDB + SwissProt + FIR + PRF) using a homology search strategy (Altschul et al., 1997). A search strategy based on a position-specific score matrix (PSSM), the psi-BLAST search, was applied to find sequences distantly related to CHH. We adopted the CHH sequence from *Macrobrachium rosenbergii* (Mar-CHH, GenBank accession number AF219382) as the initial query seed. Sequences with *E* values below 0.005 in the hitting list were selected to generate a PSSM for the next run of the psi-BLAST search. The iteration of psi-BLAST was stopped when the search result was converged—that is, no new sequence was added to the hitting list. Sequence records were downloaded to a local client in FASTA format and parsed into a standard database format with PERL (Version 5.6). The relational database was built in MySQL for sequence management. For further functional analysis, redundant sequences and those with little variation were excluded.

The DNA sequences of *Chh*-related genes and their cDNAs were collected by keyword search on GenBank through NCBI entrez interface (available at <http://www.ncbi.nlm.nih.gov/entrez/>).

In Silico Prohormone Processing. Hormone precursor sequences deduced from complementary DNA sequences were further segregated into signal peptide, precursor-related peptide, and mature hormone segments according to the results from signal peptide prediction, sequence alignments, and descriptions from feature tables in sequence flat files or from the related literature. We used SignalP (<http://www.cbs.dtu.dk/services/SignalP-2.0/>; Nielsen et al., 1999) to predict the signal peptide cleavage site. The prediction results were accepted only when an identical cleavage site was found in both artificial

neural networks (NN) and hidden Markov model (HMM) algorithms. The boundary between the precursor-related peptide and mature peptide was determined by information in the affiliated feature table or by sequence alignment of precursors and reported mature peptide sequences. When duplicate sequences from a species were reported in different databases, or by different authors, on different dates, with different methods, we preferred the longest record or the one deduced from full-length cDNA cloning data. Other processing information in the sequence documents was checked manually.

Sequence Alignment and Phylogenetic Analysis. Multiple sequence alignment (MSA) was performed on CLUSTAL X 1.83 (Thompson et al., 1997). The phylogenetic analysis was performed on MEGA 2, Version 2.1 (Kumar et al., 2001) using the preceding MSA results. A genetic distance tree was built by the minimum evolution (ME) method. The estimation of genetic distance is based on the Kumar method and Poisson correction model. The confidence values of the branching pattern were tested on 500 bootstrap replicates.

The exon-intron organization of *Chh*-related gene was determined by comparing the cDNA sequences with the genomic sequence through NCBI Spidey server (<http://www.ncbi.nlm.nih.gov/IEB/Research/Ostell/Spidey/index.html>; Wheelan et al., 2001).

Motif Analysis. The motifs of CHH/MIH/GIH family sequences were discriminated by MEME 3.0, in which the expectation-maximization algorithm is used to discover conserved regions, or motifs, in a data set of protein sequences (Bailey and Gribskov, 1998). Parameters were manipulated for the optimal setting in motif discovery. In this study we used the parameters "maximum width" setting of 20 amino acids, "maximum number of motifs" of 10, "occurrence of motif" of any number. Motif models in MEME output were assigned in ranking order. The algorithm for motif discovery was stopped when *E* was less than 0.0001.

The solution structure of MIH (PDB code 1JOT) from the Kuruma prawn *Marsupenaeus japonicus* and the simulation of a CHH isoform (Swissprot accession number Q94676) from the same species were reported (Katayama et al., 2003). The simulated CHH tertiary structure was rebuilt on the SWISS-MODEL server (<http://www.expasy.org/swissmod/SWISS-MODEL.html>). These 3-dimensional structures were optimized and colored according to the motifs found in this study with Swiss-pdb viewer 3.7 (<http://www.expasy.org/spdbv/>).

Results and Discussion

Consolidation on the CHH Family Sequence Data Set. CHH-related peptide sequences were collected from public sequence databases by homology search. There were 124 records from 34 unique species in the first BLASTp result, including 68 sequences highly related to Mar-CHH. Fifty new sequences were added to the hitting list of the psi-BLAST iteration 2. The search result was converged at iteration 3, and sequences in the hitting list were collected as the raw data set. Precursor sequences were processed as described in "Materials and Methods," and repetitive records were manually purged. The most-complete and nonredundant sequence information was therefore preserved in the working data set.

Sequence LMWP (prf accession number 2017320A), the abbreviation of low molecular weight peptide of the black widow spider *Latrodectus mactans*, was copurified with α -latrotoxin and reported to be a CHH-related sequence (Kiyatkin et al., 1992). It was not included in the psi-BLAST reports and was only found by a keyword search.

All peptide sequences and their abbreviations used in this study are listed in Table 1. In brief, 61 sequences were collected as our CHH family data set for further analysis, including 50 crustacean sequences, 9 insect sequences, with 3 novel CHH-related sequences discovered in this study (described later), one sequence from the nematode *Caenorhabditis elegans* cDNA clone, and the CHH-related sequence LMWP from the black widow spider.

Exon-Intron Organization of CHH Family Genes. There are 17 *Chh*-related gene sequences from 9 species available in public sequence databases. These genes were dissected to distinguish the exons and introns. Two types of genes in different exon-intron organizations are found (Figure 1). Type I genes are composed of 4 exons and 3 introns; for example, the *Chh* genes of *Carcinus maenas*, *Macrobrachium rosenbergii*, *Procambarus clarkii* (Figure 1, A), and the putative *Itp* gene of *Drosophila melanogaster*. Type II genes contain 3 exons and 2 introns, such as the *Chh*-like gene 43-1, 43-2, 43-3, and 43-4 of *Metapenaeus ensis*, *Moih-1*, *Moih-2*, and *Mih* genes of *Cancer pagurus* (Figure 1, B).

Exon I of type I genes contains the 5' untranslated region (5'-UTR) and a very short segment of the signal peptide. Exon II encodes a large segment of the prepropeptide, including the major part of the signal peptide, the precursor-related peptide, and the N-terminal half of the mature peptide fragment (e.g., residues 1 to 40 of Mar-CHH). Most of the type I genes were reported to have 2 transcribing forms:

Table 1. CHH family Sequences Data Set

<i>Sequence identifier</i>	<i>gi number</i>	<i>Database accession no.</i>	<i>Species</i>	<i>Remarks</i>
Arthropod: Crustacean				
Arv-CHH	399248	sp P30814	<i>Armadillidium vulgare</i>	Isopod
Arv-GIH	34924881	sp P83627	<i>Armadillidium vulgare</i>	Isopod
Byt-CHH	13506777	gb AAK28329.1	<i>Bythograea thermydron</i>	
Cam-CHH-SG	116299	sp P14944	<i>Carcinus maenas</i>	
Cam-CHH-PO	11093821	gb AAG29432.1	<i>Carcinus maenas</i>	
Cam-MIH	467733	emb CAA53591.1	<i>Carcinus maenas</i>	
Cap-CHH	3287747	sp P81032	<i>Cancer pagurus</i>	
Cap-MIH	1588530	prf 2208452A	<i>Cancer pagurus</i>	
Cap-MOIH	9931078	emb CAC05347.1	<i>Cancer pagurus</i>	
Cas-MIH	882250	gb AAA69029.1	<i>Callinectes sapidus</i>	
Chf-MIH	3776532	gb AAC64785.1	<i>Charybdis feriatus</i>	
Ers-MIH-like	18001023	gb AAL55256.1	<i>Eriocheir sinensis</i>	
Fec-MIH-like	25167482	gb AAL55258.4	<i>Fenneropenaeus chinensis</i>	
Hoa-CHH-A	930026	emb CAA38611.1	<i>Homarus americanus</i>	
Hoa-CHH-B	2497919	sp Q25154	<i>Homarus americanus</i>	
Hoa-GIH	246381	gb AAB21582.1	<i>Homarus americanus</i>	
Jal-CHH	5921749	sp P56687	<i>Jasus lalandii</i>	
Jal-MIH	34922370	sp P83220	<i>Jasus lalandii</i>	
Lie-MOIH	7447228	pir JC5628	<i>Libinia emarginata</i>	
Lis-MIH-like	18001025	gb AAL55257.1	<i>Litopenaeus stylirostris</i>	
Liv-CHH	1483572	emb CAA68067.1	<i>Litopenaeus vannamei</i>	
Liv-MIH-1	27261759	gb AAN86056	<i>Litopenaeus vannamei</i>	
Liv-MIH-2	27261761	gb AAN86057	<i>Litopenaeus vannamei</i>	
Maj-SGP I	5921746	sp O15980	<i>Marsupenaeus japonicus</i>	
Maj-SGP II	12585208	sp Q9U5D2	<i>Marsupenaeus japonicus</i>	
Maj-SGP III	1563715	gb BAA13481	<i>Marsupenaeus japonicus</i>	Maj-CHH
Maj-SGP IV	6166561	sp P55847	<i>Marsupenaeus japonicus</i>	Maj-MIH
Maj-SGP V	5921747	sp O15981	<i>Marsupenaeus japonicus</i>	
Maj-SGP VI	8928016	sp P81700	<i>Marsupenaeus japonicus</i>	
Maj-SGP VII	5921748	sp O15982	<i>Marsupenaeus japonicus</i>	
Mal-CHH	3639071	gb AAC36310.1	<i>Macrobrachium lanchesteri</i>	
Mar-CHH	17646172	gb AAL40915.1	<i>Macrobrachium rosenbergii</i>	
Mar-CHH-L	17646173	gb AAL40916.1	<i>Macrobrachium rosenbergii</i>	
Mar-SGP-A	17226814	gb AAL37948.1	<i>Macrobrachium rosenbergii</i>	
Mar-SGP-B	17226816	gb AAL37949.1	<i>Macrobrachium rosenbergii</i>	
Mee-CHH	7453857	gb AAF63028.1	<i>Metapenaeus ensis</i>	
Mee-MIH-A	3335637	gb AAC27452.1	<i>Metapenaeus ensis</i>	
Mee-MIH-B	17026322	gb AAL33882.1	<i>Metapenaeus ensis</i>	
Nen-GIH	14276186	gb AAK58133.1	<i>Nephrops norvegicus</i>	
Orl-CHH	541789	emb CAA56674.1	<i>Orconectes limosus</i>	
Pem-CHH	3986738	gb AAC84143.1	<i>Penaeus monodon</i>	
Prb-CHH	2497924	sp P55845	<i>Procambarus bouvieri</i>	
Prb-MIH	2062411	gb AAB53371.1	<i>Procambarus bouvieri</i>	
Prc-CHH	6683041	dbj BAA89003.1	<i>Procambarus clarkii</i>	
Prc-CHH-L	18766956	gb AAL79192.1	<i>Procambarus clarkii</i>	
Prc-MIH	2497929	sp P55848	<i>Procambarus clarkii</i>	
Pem-SGP C1	164163781	gb BAB70610.1	<i>Penaeus monodon</i>	
Pem-SGP C2	16151642	gb BAB69830.1	<i>Penaeus monodon</i>	
Scs-MIH	19918930	gb AAL99355.1	<i>Scylla serrata</i>	
Trc-MIH-like	18001029	gb AAL55259.1	<i>Trachypenaeus curvirostris</i>	
Arthropod: Noncrustacean				
Ang-ITP-L	30176617 ^a	gb EAA09451.2 ^a	<i>Anopheles gambiae</i> str. PEST	
Ang-ITP	- _b		<i>Anopheles gambiae</i> str. PEST	
Bom-CHH-like	13431368	sp Q9NL55	<i>Bombyx mori</i>	
Drm-ITP-L	16648484	gb AAL25507.1	<i>Drosophila melanogaster</i>	
Drm-ITP	- _b		<i>Drosophila melanogaster</i>	
LMWP	1351906	sp P49125	<i>Latrodectus mactans</i>	
Lom-ITP	4433632	gb AAD20820.1	<i>Locusta migratoria</i>	
Lom-ITP-L	4433634	gb AAD20821.1	<i>Locusta migratoria</i>	

(continued)

Table 1. Continued

Sequence identifier	gi number	Database accession no.	Species	Remarks
Scg-ITP	1244522	gb AAB16822.1	<i>Schistocerca gregaria</i>	
Scg-ITP-L	1244524	gb AAB16823.1	<i>Schistocerca gregaria</i>	
Nonarthropod				
Cae-CHH-like	17544482	ref NP_501 985.1	<i>Caenorhabditis elegans</i>	

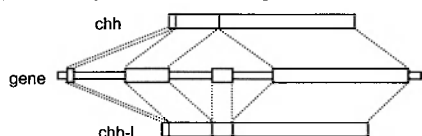
^aSequences used in this study are modified and deduced from the original one (gi number 21297306) due to inadequate data for open reading frame prediction.

^bThese sequences are predicted by the rules of alternative splicing.

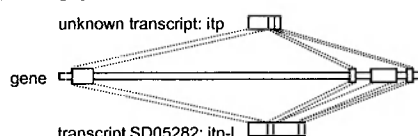
A. CHH related gene structure

• the four-exon gene

(1) the Chh gene of *M. rosenbergii*



(2) the Itp gene of *D. melanogaster*

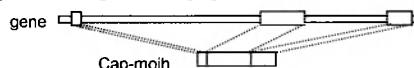


• the three-exon gene

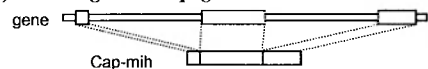
(1) a Chh gene of *M. ensis*



(2) the Moih gene of *C. pagurus*



(3) the Mih gene of *C. pagurus*



B. Amino acid sequence alignment.

• peptide sequences of some four-exon gene transcripts

Mar-CHH	AILDQSKGIFDRELFPKKLDRIEDYNYLYRKPYPVAIDREGYQNLVFRQIQDLQLMDQLDEYANAVQIVGK	exon II	exon III
Mar-CHH-L	AILDQSKGIFDRELFPKKLDRIEDYNYLYRKPYPVAIDREGYQNLVFRQIQDLQLMDQLDEYANAVQIVGK	exon II	exon IV
Drm-ITP	SNFFDLEKGIENKTMFFRLDRIEDYNYLYRKPYPVAIDREGYQNLVFRQIQDLQLMDQLDEYANAVQIVGK	exon II	exon IV
Drm-ITP-L	SNFFDLEKGIENKTMFFRLDRIEDYNYLYRKPYPVAIDREGYQNLVFRQIQDLQLMDQLDEYANAVQIVGK	exon II	exon IV

• peptide sequences of some putative four-exon gene transcripts

Scg-ITP	SFFDIOKGVYDKSIFARLDRIEDYNYLYRKPYPVAIDREGYQNLVFRQIQDLQLMDQLDEYANAVQIVGK	exon II	exon III
Scg-ITP-L	SFFDIOKGVYDKSIFARLDRIEDYNYLYRKPYPVAIDREGYQNLVFRQIQDLQLMDQLDEYANAVQIVGK	exon II	exon IV
Ang-ITP	SSFFDIEKGVYDKSIFARLDRIEDYNYLYRKPYPVAIDREGYQNLVFRQIQDLQLMDQLDEYANAVQIVGK	exon II	exon III
Ang-ITP-L	SSFFDIEKGVYDKSIFARLDRIEDYNYLYRKPYPVAIDREGYQNLVFRQIQDLQLMDQLDEYANAVQIVGK	exon II	exon IV

• peptide sequences of some three-exon gene transcripts

Mee-CHH	SLFDPSITGVFDR-ELLGRLNRVEDYNYLYRKPYPVAIDREGYQNLVFRQIQDLQLMDQLDEYANAVQIVGK	exon II	exon III
Cap-MIH	RVINDLGNLGNRDLYKKVWIEDYNYLYRKPYPVAIDREGYQNLVFRQIQDLQLMDQLDEYANAVQIVGK	exon II	exon III
Cap-MOIH	RRINNDQNFIGNRAMYKVDWIEDYNYLYRKPYPVAIDREGYQNLVFRQIQDLQLMDQLDEYANAVQIVGK	exon II	exon III

Fig. 1. Chh-related gene structures. A: Gene structure of some CHH family sequences. B: Alignment of some putative mature peptides of CHH family genes. The common segments of each spliced pair are boxed. The cysteine residues are highlighted. The exon-exon junction boundaries are indicated by arrows. A broken line in Mee-CHH indicates the gap introduced in the alignment.

one contained exon III in the coding region, while another bypassed it. In fact, the peptide from the latter transcribing form (exon I-II-IV) is the reported hyperglycemic hormone in these species, and the transcribing form with exon III (exon I-II-III-IV) was considered as a novel isoform (Chen et al., 2003; Dirksen et al., 2001). In the case of the transcript of exon I-II-IV, exon IV contributes the latter half segment of the peptide (e.g., residue 41 to the C-terminal end of Mar-CHH) and the 3'-UTS. When exon III is adjoined to exon II, for example, the transcript of exon I-II-III-IV, the C-terminal peptide segment will be encoded by exon III rather than by exon IV. However, the two cysteine residues in this segment are conserved in both cases (Figure 1, B).

Exons I and II of type II genes contain the 5'-UTR, the signal peptide, and the N-terminal half of the mature peptide. Exon III of type II genes encodes the C-terminal segment and the 3'-UTR. Although these two types of CHH family genes were found to differ in exon numbers, a common design at the exon-exon junction within the mature peptide segments was observed (Figure 1, B). The reading frame of a mature peptide segment is separated by a phase 2 intron (exon-intron boundary after the second base of a codon) in both cases. The residue at the breaking point is located right after the fourth conserved cysteine; only arginine or lysine is found at this point (Figure 2).

Some sequences have characters resembling type I genes, but evidence is not yet conclusive. The *Chh* gene of *Macrobrachium lanchesteri* (*Mal-Chh* gene, GenBank accession number AAC36310.1) was annotated to have three exons. When we compared the *Chh* gene of *Macrobrachium rosenbergii* (*Mar-Chh* gene) with the *Mal-Chh* gene, the 4 exons and 3 introns matched their counterparts with high identity scores; therefore, we proposed that the *M. lanchesteri Chh* gene may have exon III (Chen et al., 2003). In another case, ITP and ITP-L from *Locusta migratoria* and *Schistocerca gregaria* were considered as alternative transcribing products from their genes (Audsley et al., 1992; Macins et al., 1999). The pattern of open reading frame switching and the exon-exon junction within the mature peptide reading frame are identical to the 4-exon *Chh* gene (Figure 1, B). Although the sequences of locust *Itp* genes are not yet available, the putative insect *Itp* gene from an other species, *D. melanogaster*, is in a 4-exon structure, from which the putative alternative spliced transcripts were discovered in this study (Figure 1, B). Besides, two other sequences from *L. vannedmei*, Liv-MIH-1 and Liv-MIH-2, are annotated as spliced products. They are similar to the ITP/ITP-L from locusts and CHH/CHH-L from crustaceans,

suggesting that these alternative spliced transcripts may originate from a 4-exon gene as we described above.

The *chh-like* cDNA of *Caenorhabditis elegans* (ZC168.2, GenBank accession number NP_501985.1) was compared to its genome. The alignment was divided into two segments, but the cDNA sequence was incomplete because of the missing 5'- and 3'-UTRs. Consequently, the gene organization is still inconclusive; however, a complete intron with boundaries is well defined, and the exon-exon junction within the mature peptide is disclosed. The trinucleotide codon of the residue after the fourth conserved cysteine, AGA for the arginine in this case, is also interrupted by a phase 2 intron, as also observed in the other *Chh*-related genes.

Data-mining on Insect Sequences. Several insect peptide sequences are accessible in our data set on the basis of sequence homology, but their biological functions have not been presented sufficiently. The Lom-ITP from *L. migratoria* and Scg-ITP from *S. gregaria* were characterized on their function of ion transportation, but the biological roles of ITP-L peptides from the same species were unknown (Audsley et al., 1992; Macins et al., 1999). The CHH-like sequence from the silk moth *Bombyx mori* (Bom-CHH-like) was deduced from the cDNA cloned with degenerated primers designed from the consensus amino acid sequence among crustacean CHHs and ITPs (Endo et al., 2000). Bioassay data are unavailable for this sequence. A peptide from the fruit fly *D. melanogaster* with CHH/MIH/GIH family characters was reported from the sL2 cell cDNA library clone SD05282 (Drm-ITP-like, GenBank accession number AAL25507.1). Four exons on the putative *Itp* gene were defined (CG13586; FlyBase ID, FBgn0035023) (Figure 1, A). Only one cDNA sequence has been reported; however, we wondered if the same transcribing rules as for the other type I genes, the 4-exon *Chh* genes, would be applicable to it.

A novel peptide sequence was deduced from the joined sequence of exon I-II-IV (Figure 2, Drm-ITP). Two of the 6 positional conserved cysteine residues in the C-terminal segment are well preserved. It is interesting that a GRK sequence found in its C terminus is commonly seen in the CHH family peptides and can be recognized as an endopeptidase processing site. On the presumption of a common posttranslational processing fate for the CHH subfamily I members, it may encode a 145-residues precursor peptide and finally become a mature peptide of 73 residues with an amidated C-terminus. Accordingly, we presume that the putative peptide of exon I-II-IV transcript is homologous to the insect

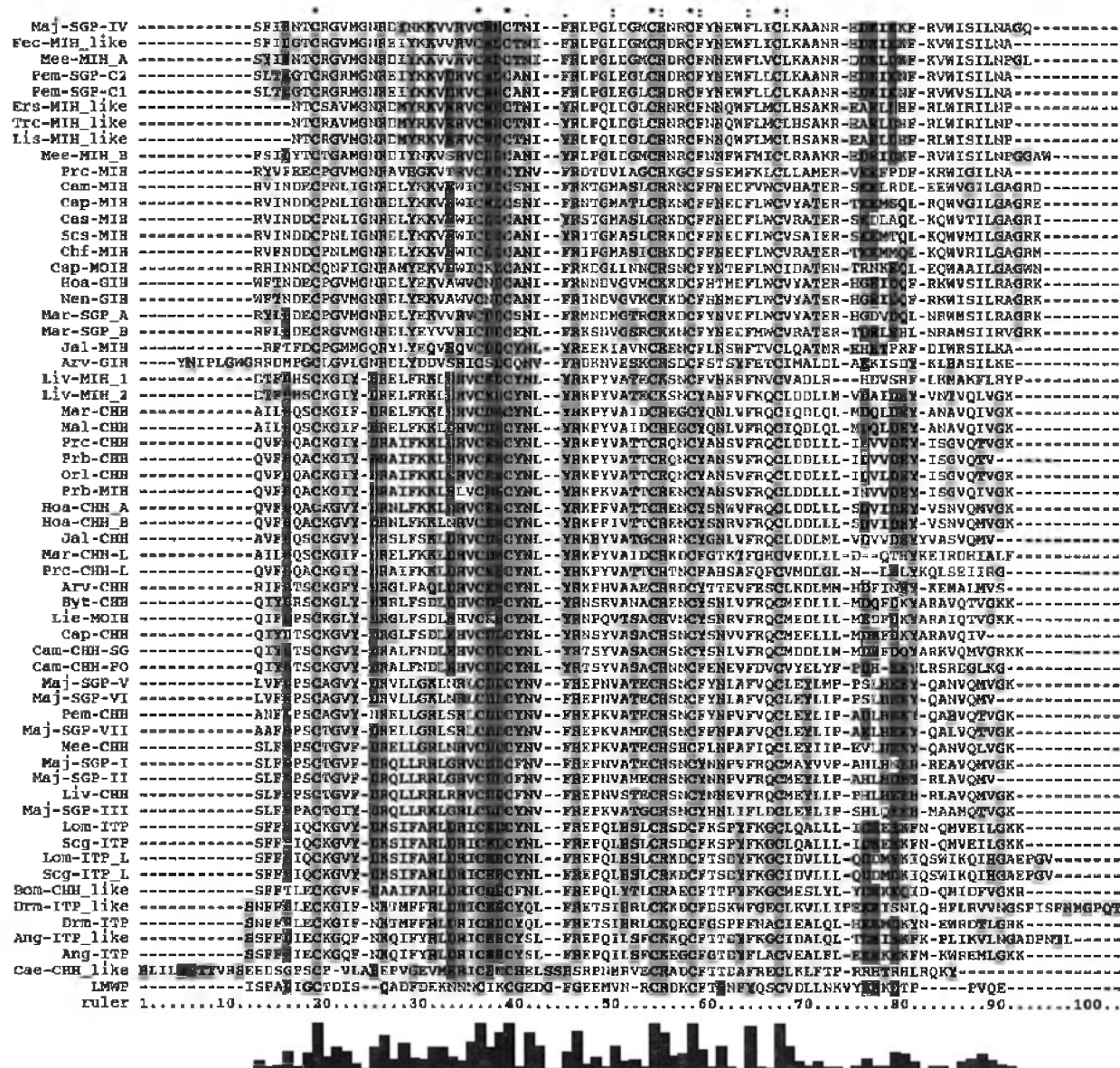


Fig. 2. Multiple-sequence alignment of CHH-related sequences by CLUSTAL X 1.83. The symbol at the top represents the highly conserved position in the alignment table. An asterisk (*) indicates that all residues in a column are identical. A colon (:) indicates that only closely related residues are found at this position. The bar chart at the bottom reveals the score evaluating consensus status among sequences in the alignment result.

ITP, named Drm-ITP, and the exon I-II-III-IV transcript (equivalent to the SD05282 cDNA sequence), named Drm-ITP-L.

Another sequence from *Anopheles gambiae* (Ang-ITP-like) was annotated as a CHH/MIH/GIH neuropeptide without the expression data and contains only 4 cysteines in the predicted peptide sequence (Riehle et al., 2002). Through Ensembl Genome Browser (<http://www.ensembl.org>), the CHH homologue sequence from the anopheles genome was retrieved by homologue search strategies. When we started the TBLASTn query with the Drm-

ITP-L peptide sequence, several aligned segments were obtained. One alignment contained the first half of Drm-ITP-L. It matched a peptide sequence with 4 cysteine residues that was enclosed in the predicted annotated reading frame. Two other segments on the anopheles genome were aligned to the latter half of Drm-ITP-L. Both of them contained 2 cysteine residues that aligned to the fifth and the sixth cysteine residues of the query sequence with a perfect spacing pattern.

The alignment results revealed the possibility of inaccurate prediction. These aligned segments were

all located on chromosome 2L within a restricted range (about 4 kb) on the same DNA strand. The segment aligned to the first half of Drm-ITP-L preceded the other 2 segments. Moreover, the boundaries of the aligned segments were also nearly the same to the exon-exon junction within the mature peptide sequence as in the other CHH family members. We therefore attempted to retrieve the putative open reading frames from a 12-kb genome segment (sequence 2L.8306909-8319589), in which the annotated gene is enclosed. After inspection of the sequence context of the predicted exons, we found deduced peptide sequences of some exon candidates could be aligned to those from exons II, III, and IV of Drm-ITP with high scores. Therefore, the 3 exon candidates were presumed to be the counterparts of Drm-Itpr exons II, III, and IV. The exon I homologue was not detected from the genome. From the rules of exon usage observed in the transcripts of the 4-exon *Chh* genes, 2 putative precursor peptides with a dibasic residue cleavage site were deduced. The coding regions of putative exon II-III-IV [gb|AAAB01008900 region <356740..356957, 359188..>359308] was presumed as the partial prohormone sequence of Ang-ITP-L, and the prohormone of Ang-ITP was from the coding region of putative exon II-IV [gb|AAAB01008900 region <356740..356957,360354..>360459]. The putative dibasic cleavage site was further identified when the 2 prohormone sequences were aligned to the other insect sequences. The 2 novel presumptive peptides found by this approach contained all 6 positional conserved cysteine residues, and a possible C-terminal processing sequence, GKK, was found in the peptide from the short transcript (exon II-IV) (Figure 1, B).

Sub grouping of CHH Family Members. The CHH family was divided into 2 subfamilies according to sequence similarities (Keller, 1992; de Kleijn and Van Herp, 1995; Lacombe et al., 1999; Bocking et al., 2001). In this study we attempted to reclassify the CHH-related sequences on the updated data set including the recently discovered ITPs, MOIHs, and several undefined sequences, by methods of multiple sequence alignment and phylogenetic analysis. As shown in Figure 2, highly conserved sites including the 6 positional conserved cysteines were marked on the top of the alignment table. Meanwhile, the conservation of each column was calculated in the alignment quality score plotted beneath. Most of the conserved residues were concentrated around the cysteines, and sequences after the last cysteine were more divergent from each other.

The dendrogram of CHH-related sequences was built on MEGA 2 with the multiple sequence

alignment results from CLUSTAL X (Figure 3). All sequences except LMWP and Cae-CHH-like were segregated into 3 major groups. The insect sequences were clustered in the ITP subgroup. These sequences are closely related to each other and may perform similar biological functions *in vivo*. All CHHs, Maj-SGP-I, II, III, V, VI, and VII, Liv-MIH-1, Liv-MIH-2, Prb-MIH, and Lie-MOIH, were found in the second subgroup, namely, the CHH subgroup. The last subgroup, the MIH/GIH subgroup, consisted of most of the MIHs, GIHs, Cap-MOIH, Mar-SGP-A, Mar-SGP-B, Maj-SGP-IV, Pem-SGP-C1, and Pem-SGP-C2.

Previously observed sequence characters are supported by this approach. For example, a precursor-related peptide and a KR or RR dibasic residue processing site were found in precursor sequences of the CHH and ITP subgroups, while the signal peptide of most MIH and GIH prohormones were found to precede the mature peptide directly. In addition, a distinctive difference between these 2 subfamilies in the mature peptide region was present in the alignment table. A deletion of amino acid residue was found in the sequences of the CHH and ITP subgroups at the fifth position after the first conserved cysteine residue, where a gap is denoted, contrasting to the G at the same position of the MIH/GIH subgroup sequences. Furthermore, we found another distinctive site in the primary structure level: the fourth position preceding the second conserved cysteine was L in the CHH and ITP subgroups, but it was substituted by V in the MIH/GIH subgroup. According to the general principles of the CHH subfamily classification (Keller, 1992; de Kleijn and Van Herp, 1995; Lacombe et al., 1999; Bocking et al., 2001), the MIH/GIH subgroup is equivalent to CHH subfamily II, or the MIH/GIH subfamily, and the previously described ITP along with CHH subgroups comprised the CHH subfamily I (Figure 3, B). Since all ITP sequences are classified into the CHH subgroup, we suggest the CHH/ITP subfamily is synonymous with CHH subfamily I.

Two MOIH sequences were classified into different subfamilies: MOIH of *Cancer pagurus* (Cap-MOIH) was in the MIH/GIH subfamily, while that of *Libinia emarginata* (Lie-MOIH) was in the CHH/ITP subfamily. In the Cap-MOIH precursor sequence, a signal peptide cleavage site was identified at GLA³⁴-R³⁵R, followed directly by its mature peptide sequence. MOIH fractionated and separated from the CHH and the MIH peaks by reverse-base HPLC was reported to inhibit MF synthesis in the mandibular organs and to have a weak inhibiting effect on ecdysteroid synthesis (Wainwright et al., 1996). However, the dual biological activity of HPLC-purified

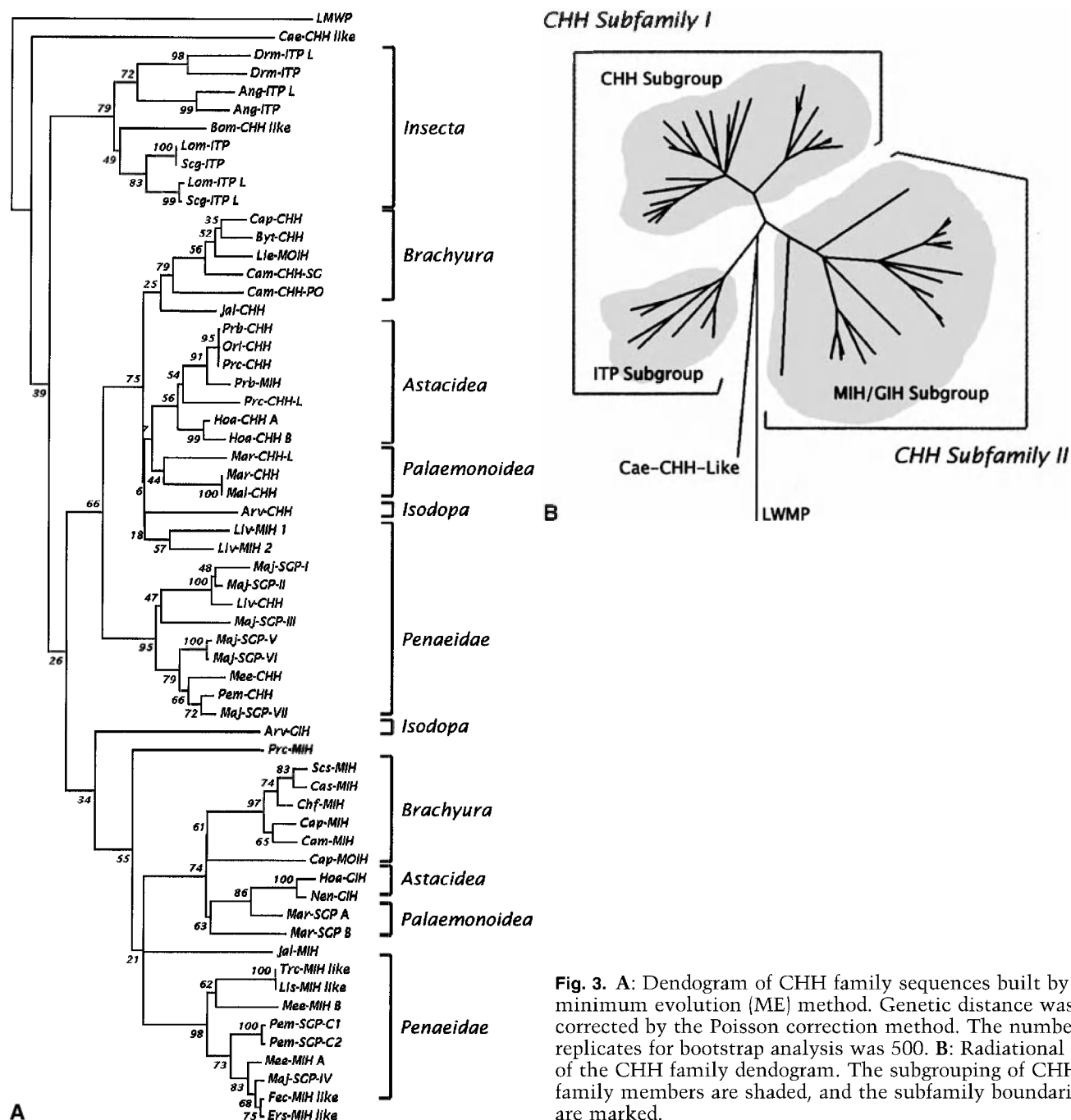


Fig. 3. A: Dendrogram of CHH family sequences built by the minimum evolution (ME) method. Genetic distance was corrected by the Poisson correction method. The number of replicates for bootstrap analysis was 500. B: Radiational tree of the CHH family dendrogram. The subgrouping of CHH family members are shaded, and the subfamily boundaries are marked.

MOIH peptide of *Libinia emarginata* was demonstrated to perform dual functions, inhibition of MF synthesis as MOIH and a hyperglycemic effect as CHH (Liu et al., 1997). A precursor-related peptide with the dibasic cleavage signal was found in the prepropeptide sequence of Lie-MOIH, whose sequence resembled those of the CHH subfamily sequences. Since reported information on MOIHs was scant, the controversy is cannot be clarified here.

Inconsistency between the classification of molecular characters and nomenclature was also

observed for the MIH from *P. bouvieri* (Prb-MIH) and 2 MIH sequences from *Litopenaeus vannamei* (Liv-MIH-1 and Liv-MIH-2). Prb-MIH was proved not to have a hyperglycemic effect. The mature peptide sequence characters were closer to CHH, while its prohormone sequence was unavailable. Partial prohormone sequences of Liv-MIH-1 and Liv-MIH-2 were submitted to GenBank directly without reference to their characterization, identification, or physiological effect. Bocking et al. (2001) summarized the complicated situation on the pleiotropic

potencies and multiple forms of CHH family peptides. Reexamination of biological potency, characterizing the hormone receptors for specific isoforms, deciphering the complete precursor sequences, or other experimental approaches for these molecules would be necessary to moderate the discrepancy.

CHH-like Peptides from *Caenorhabditis elegans* and *Latrodectus mactans*. The expression of CHH-like gene in a nematode, *C. elegans* was observed. CHH-like sequence from *C. elegans* (Cae-CHH-like) was predicted from its genome sequence. It was proven to be expressed at the dauer larvae stage in the serial analysis of gene expression (SAGE) data (Jones et al., 2003). Since the CHH-related sequence has never been reported from species other than arthropods, the homologous relationship of Cae-CHH-like to the CHH family members is especially notable. First, we identified a signal peptide cleavage site at SSA²⁵-R²⁶L. However, no dibasic residue cleavage signal can be identified between the predicted signal peptide cleavage site and the first conserved cysteine residue. It is much more diverged from the other sequences in our data set; thus, it caused several extra gaps in the multiple sequence alignment result (Figure 2). The *Cae-Chh-like* gene shares the common exon-exon junction pattern within the mature peptide segment, but the gene boundary and the exon-intron organization are not conclusive for the incompleteness of the cDNA sequence.

Annotation of LMWP as a CHH-like sequence in the original literature was based on the number and spacing pattern of the cysteine residues in the primary sequence, similar to a CHH family member (Kiyatkin et al., 1992). When we started a BLAST search with LMWP, several sequences were listed in the report, but none of them were relevant to the CHH family. The search revealed that the physicochemical characters of the residues in the LMWP sequence differed greatly from the CHH family members. The low homologous relationship between LMWP and the CHH family members was also reflected in the topology of the dendrogram (Figure 3). LMWP left the tree earliest, even before the branching point of the nonarthropod sequence Cae-CHH-like. Therefore, we concluded that the Cae-CHH-like is a CHH family member, but its sequence characters are insufficient to determine its subfamily classification owing to its divergency and the uncertainty of the sequence context; LMWP should be excluded from the CHH family.

Motif Analysis. Mature crustacean peptide sequences in the data set were subjected to motif analysis. To express the residue position for each

motif succinctly, we adopted Mar-CHH as a representative sequence in the following description. When we used MEME 3.0 to identify protein motifs of the family, a large motif starting from 9 to 61 was detected on the default settings. All 6 conserved cysteine residues were enclosed in this large motif. By manipulating parameters to optimal conditions, the large motif was split up. Eight motifs were found and named by ranking order. The motif configurations in each sequence are shown in Figure 4(A). The core motifs, motif 1 (21-40) and motif 3 (42-61), were found in all submitted sequences, but motifs in the N-terminal and the C-terminal regions of the CHHs and MIHs/GIHs were different. Generally, the CHH subfamily I members were composed of 4 motifs, arranged in the order 2-1-3-5, and the motif configuration of subfamily II was in the order 4-1-3-6. The motif preference of the 2 subfamilies may reveal the critical roles of these motifs in functional specificity.

Variations of the motif configurations were observed. Motif 5 was absent in the C-terminal segment of Mar-CHH-L, Cam-CHH-PO, Prc-CHH-L, and Liv-MIH-1. These sequences were previously reported as spliced products (Chen et al., 2003; Dirksen et al., 2001). Other non-CHH sequences in subfamily I, such as MIH from *P. bouvieri* and MOIH from *L. emarginata*, were found in the typical 2-1-3-5 configuration. Differences in the configurations of some MIH/GIH subfamily sequences in the N-terminal or C-terminal segments are due to the sequences' incompleteness; motif 7 is overlapped by motif 6 and should be considered as a redundant motif.

The discovered motifs were submitted to MAST (Motif Alignment and Search Tool, <http://memes.sdsc.edu/meme/website/mast.html>) to search for homologues against a nonredundant database. The CHHs and MIHs/GIHs were in typical motif configurations as described above. Meanwhile, the ITPs and other CHH-related peptides from insects were found to have the core motifs (motifs 1 and 3) and the N-terminal motif of CHHs (motif 2). Motif 3 did not appear in the predicted CHH-like sequence from *Anopheles gambiae*. This is consistent with our previous suspicion that predictions of the peptide sequence after the fourth conserved cysteine were incorrect. Nevertheless, Cae-CHH-like and LMWP were not found by this attempt.

When the motifs detected in our study were compared with those reported in public databases, the coverages of motifs were similar. The PRINT entry PR00550, mentioned in INTERPRO entry IPR001166 (Crustacean neurohormone CHH/MIH/GIH) as a family signature, defines 3 motifs of the CHH-related sequences. Motifs 1-2-3 cover a 42-

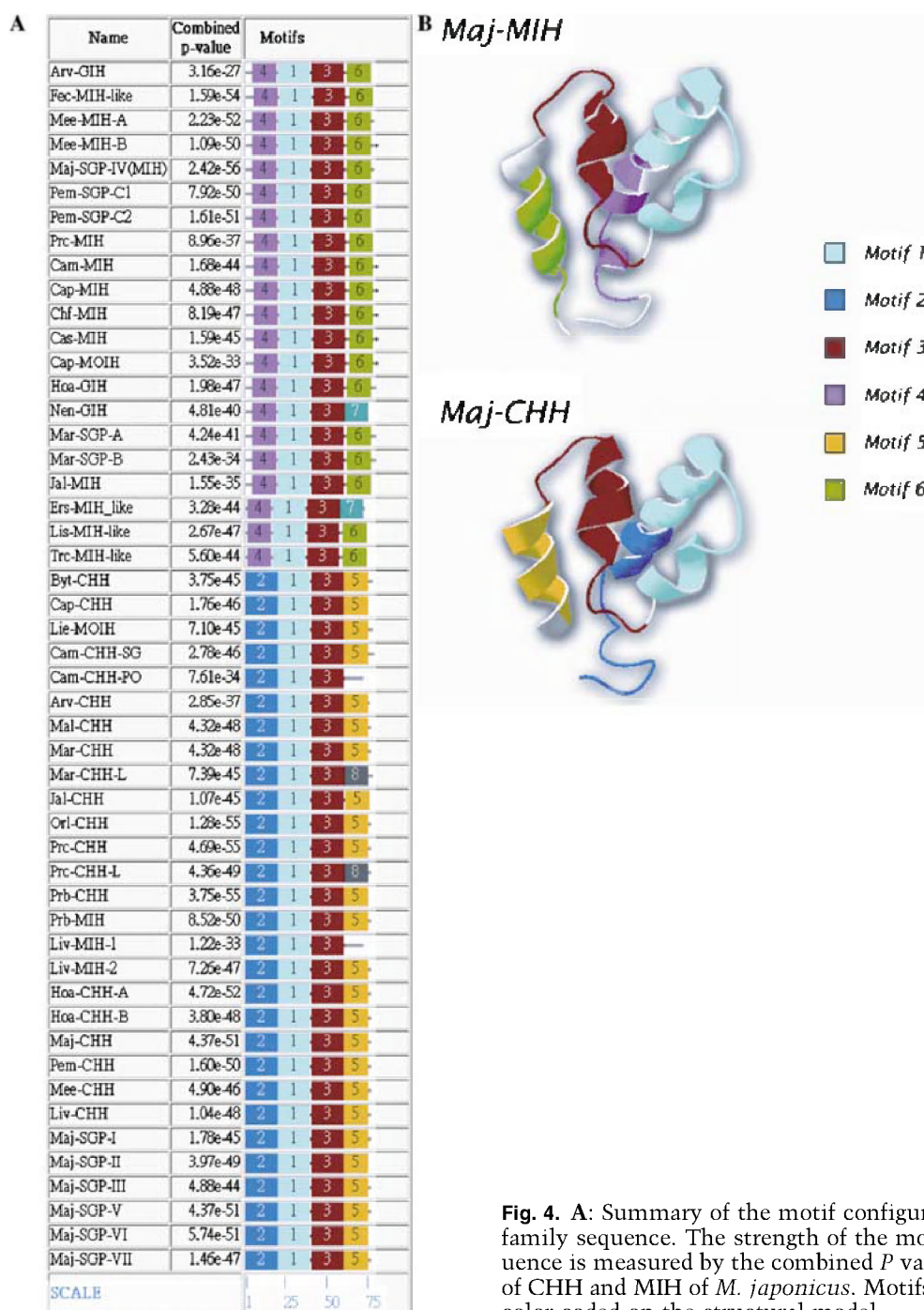


Fig. 4. **A:** Summary of the motif configuration of the crustacean CHH family sequence. The strength of the motif configuration in a given sequence is measured by the combined *P* value. **B:** Tertiary structure models of CHH and MIH of *M. japonicus*. Motifs discovered by this approach are color-coded on the structural model.

residue segment that shifted 10 residues leftward to the core motifs (motifs 1 and 3) of this report. IPR000346 (hyperglycemic hormone type 1, PR00548) and IPR001262 (hyperglycemic hormone type 2, PR00549) described the motif signature of CHH subfamily I and II respectively. Each of them contains 4 motifs, and the motif allocations are compatible with the 4 motifs defined in this study. However, the sequence contexts of these 2 motif

definitions did not cover all of the positional conserved cysteine residues in the signature, which are the most critical characters of CHH family peptides. The 2 motif signatures were difficult to compare with each other from a functional view point, for the motifs they described were not aligned by position and failed to describe the common segment of the 2 subfamilies. In contrast, we defined the common character of CHH-related sequences by the core

motifs (motifs 1 and 3). The core motifs cover 5 of the 6 conserved cysteine residues. The motif signatures for the 2 subfamilies can be easily described as the basic component (the core motifs) plus the subfamily-specific N-terminal motifs and C-terminal motifs. The motif configurations of the 2 subfamilies also imply the significance of the N-terminal and C-terminal motifs in peptide functions.

Molecular Anatomy of the CHH Family. The first 3-dimensional structural model of MIH of *M. japonicus* (Maj-MIH) was resolved by Katayama et al. (2003). Spatial allocations of the motifs identified in this study were attempted (Figure 4, B, lower plate). The core motifs formed the general shape of the tertiary structure of CHH family peptides by which 5 of the 6 cysteine residues were enclosed. The remaining cysteine residue was within the N-terminal motifs. It linked to the fifth cysteine residue within the core motifs, and the mobility of the N-terminal motifs was therefore restricted. The helix $\alpha 1$ in the Maj-MIH structure was considered a part of the determinants of the functional specificity (Katayama et al., 2003). We found it located within the N-terminal motif. It was composed of 3 residues and ended at the G¹² that was aligned to a gap in the CHHs. The N-terminal motifs included a coiled-coil structure and part of the helix $\alpha 2$ defined by Katayama et al. (2003), and the C-terminal motifs were in a helix structure mainly. These 2 motifs were close to each other in the 3-dimensional model. The structural model of the other CHHs is not easily built by applying the same modeling method. The sequence alignment between the target and the template may fail after the 58th residue or so in the PDB first approach mode. Nevertheless, the threading method can be used to adjust the alignment resulting from some of the poor modeling.

The secondary structure server, PHD, predicted residues in the C-terminal segment of CHHs from *M. rosenbergii*, *C. maenus*, and *P. clarkii* to be an all-helix structure (Chen et al., 2003). Each helix had medium to high confidence scores. Furthermore, the location of the predicted helix could be lined up with the MIH structure. It is therefore logical to presume a helix structure for the C-terminal motif of CHHs, while the helix structure may deviate from the position of the MIH model. In addition, the C-terminal amidation was observed in CHHs and ITPs (Nassel, 2002; Katayama et al., 2003). In vitro expression experiments with these molecules revealed the importance of the C-terminal modification to biological activity (Wang et al., 2000; Katayama et al., 2002). The effect of the modification on the structural model is however, uncertain.

The N-terminal motifs (2 and 4) and the C-terminal motifs (5 and 6) of CHHs and MIHs apparently differ from each other. These two motifs approach each other in the tertiary structure models, and the motif preference reveals the critical roles of these regions in functional specificity. It will be helpful to focus on these motifs when experimental approaches such as use of synthetic peptides, chimeric recombinant protein construction, ligand-receptor interaction, and the choice of epitopes for antibody induction are employed to study the functional specificity of members of this peptide family. In contrast, the motif configuration of GIHs was indistinguishable from that of MIHs. Thus we were unable to determine specific functional domains to regulate gonad activities. In view of molecular evolution, it is plausible that the N-terminal and C-terminal segments play important roles in hormone function.

Evolution of CHH Family Genes. Our study presents some interesting viewpoints on the evolution of CHH family genes. First, the exon-exon junction pattern within the mature peptide coding region is universal among available CHH-related genes. The trinucleotide code at the junction is separated a phase II intron. Residues found at the junction are also highly conserved; only R or K is found here. Not only arthropod sequences but also the nematode CHH-like sequence follows this rule.

Second, two types of gene organizations are observed in the family. The 4-exon *Chh*-related gene appears to be common in a wide range of crustacean species, including Astacidea, Brachyura, and Palaeomonidae, as well as in insect species. These 4-exon *Chh* genes have 2 transcripts caused by alternative splicing (Chen et al., 2003; Dirksen et al., 2001). From the rules by which 4-exon *Chh* genes make their transcripts, we discovered a potential transcript from *D. melanogaster Itp* gene. Two novel potential transcripts with CHH family characters were discovered by the same rules from *A. gambiae* genome in this study. Three *Mih/Gih* gene sequences are reported to have a 3-exon organization. In addition, *Chh* genes from *M. ensis*, reported by Gu and Chan (1998), also have 3-exon structure.

Furthermore, the existence of *Cae*-CHH-like should be considered as a clue to the long evolutionary history of the CHH-related sequences. It also suggests that the CHH-like molecules may be found in organisms other than pan-crustacean species with vital importance. The structure of the *Cae-Chh-like* gene remains to be characterized, depending on the completeness of the cDNA sequence. It may resemble the prototype of the *Chh*-related gene. Some studies on

arthropod phylogeny have revealed the close relationship between the crustacean and hexapod taxa (Shultz and Regier, 2000; Girbet et al., 2001). In a presumption of the early segregation of hexapods and crustaceans, all *Chh*, *Itp*, *Mih*, and *Gih* genes may be derived from a common gene of the ancestral species. This gene became diversified over with the evolutionary history through gene (or exon) duplication events and the accumulation of variances by mutations such as base insertion, deletion, or substitution, and finally, the relative molecules with new functions appeared and might contribute a new advantage for the fitness of species. The ITP subgroup disappeared from the dendrogram first, which implies that the segregation of hexapod taxa and crustacean taxa was complete before the gene evolution of CHHs and MIHs/GIHs began (Figure 3). Therefore, a new mechanism of molting control governing body enlargement during the whole life span was observed in crustaceans, but not in insects, which would account for the negative finding of MIH/GIH peptides in the insect species. Hence the hypothesis that MIH/GIH is unique in crustacean taxa can only be proved by knowledge of the spectrum and the sequence contexts of the CHH-related peptides in these species, or by comparative studies on the molting control mechanism from a wide range of invertebrate species.

Molecular evolution of the CHH neuro peptide family remains to be explored. With the limited availability of sequence data, we are unable to infer a model of evolution for *Chh*-related genes. It is, however, anticipated that the ontogeny of the CHH family will be better understood as new members are deciphered by cloning or through the genome projects in a wide range of species, as well as from progressive development in data-mining methods.

References

- Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25, 3389–3402
- Audsley N, McIntosh C, Phillips JE (1992) Isolation of a neuro peptide from locust corpus cardiacum which influences ileal transport. *J Exp Biol* 173, 261–274
- Bailey TL, Gribskov M (1998) Combining evidence using p-values: application to sequence homology searches. *Bioinformatics* 14, 48–54
- Bocking D, Dirksen H, Keller R (2001) The crustacean neuro peptides of the CHH/MIH/GIH family: structures and biological activities. In: *The Crustacean Nervous System*, Wiese K, ed. (New York, NY: Springer Verlag) pp 84–97
- Chang ES, Prestwich GD, Bruce MJ (1990) Amino acid sequence of a peptide with both molt-inhibiting and hyperglycemic activities in the lobster, *Homarus americanus*. *Biochem Biophys Res Commun* 171, 818–826
- Chen SH, Lin CY, Huang FL, Kuo CM (2003) Cloning of two crustacean hyperglycemic hormone isoforms from the freshwater giant prawn, *Macrobrachium rosenbergii*: evidence of alternative splicing. *Mar Biotechnol* 6, 83–94
- Dirksen H, Bocking D, Heyn U, Mandel C, Chung JS, Baggerman G, Verhaert P, Daufeldt S, Plosch T, Jaros PP, Waelkens E, Keller R, Webster SG (2001) Crustacean hyperglycaemic hormone (CHH)-like peptides and CHH-precursor-related peptides from pericardial organ neurosecretory cells in the shore crab, *Carcinus maenas*, are putatively spliced and modified products of multiple genes. *Biochem* 356, 159–170
- Endo H, Nagasawa H, Watanabe T (2000) Isolation of a cDNA encoding a CHH-family peptide from the silkworm *Bombyx mori*. *Insect Biochem Mol Biol* 30, 355–361
- Girbet G, Edgecombe GD, Wheeler WC (2001) Arthropod phylogeny based on eight molecular loci and morphology. *Nature* 413, 157–161
- Gu PL, Chan SM (1998) The shrimp hyperglycemic hormone-like neuro peptide is encoded by multiple copies of genes arranged in a cluster. *FEBS Lett* 441, 397–403
- Huberman A (2000) Shrimp endocrinology: a review. *Aquaculture* 191, 191–208
- Jones SJ, Riddle DL, Pouzyrev AT, Velculescu VE, Hillier L, Eddy SR, Stricklin SL, Baillie DL, Waterson R, Marra MA (2003) Changes in gene expression associated with developmental arrest and longevity in *Caenorhabditis elegans*. *Genome Res* 11, 1323–1324
- Katayama H, Ohira T, Aida K, Nagasawa H (2002) Significance of a carboxyl-terminal amide moiety in the folding and biological activity of crustacean hyperglycemic hormone. *Peptides* 23, 1537
- Katayama H, Nagata K, Ohira T, Yumoto F, Tanokura M, Nagasawa H (2003) The solution structure of molt-inhibiting hormone from the Kuruma prawn *Marsupenaeus japonicus*. *J Biol Chem* 278, 9620–9623
- Keller R (1992) Crustacean neuro peptides: structures, functions and comparative aspects. *Experientia* 48, 439–448
- Keller R, Kegel G, Reichwein B, Sedlmeier D, Soye Z D (1999) Biological effects of neuro hormones of the CHH/MIH/GIH peptide family in crustaceans. In: *Recent Developments in Comparative Endocrinology and Neurobiology*, Roubos SE, Vaudry H, De Loof A, eds. (Nijmegen, Netherlands: Shaker) pp 209–212
- Khayat M, Yang W, Aida K, Nagasawa H, Tietz A, Funkenstein B, Lubzens E (1998) Hyperglycaemic hormones inhibit protein and mRNA synthesis in *in vitro*-incubated ovarian fragments of the marine shrimp *Penaeus semisulcatus*. *Gen Comp Endocrinol* 110, 307–318
- Kiyatkin N, Dulubova I, Chekhovskaya I, Lipkin A, Grishin E (1992) Structure of the low molecular weight protein copurified with alpha-latrotoxin. *Toxicon* 30, 771–774

19. de Kleijn DP, van Herp F (1995) Molecular biology of neurohormone precursors in the eyestalk of Crustacea. *Comp Biochem Physiol B* 112, 573–579
20. de Kleijn DP, Janssen KP, Waddy SL, Hegeman R, Lai WY, Martens GJ, van Herp F (1998) Expression of the crustacean hyperglycaemic hormones and the gonad-inhibiting hormone during the reproductive cycle of the female American lobster *Homarus americanus*. *J Endocrinol* 156, 291–298
21. Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA 2: Molecular Evolutionary Genetics Analysis software. *Bioinformatics* 17, 1244–1245
22. Lacombe C, Greve P, Martin G (1999) Overview on the sub-grouping of the crustacean hyperglycemic hormone family. *Neuropeptides* 33, 71–80
23. Laufer H, Biggers WL, Ahl JS (1998) Stimulation of ovarian maturation in the crayfish *Procambarus clarkii* by methyl farnesoate. *Gen Comp Endocrinol* 111, 113–118
24. Laufer H, Ahl J, Rotllant G, Baclaski B (2002) Evidence that ecdysteroids and methyl farnesoate control allometric growth and differentiation in a crustacean. *Insect Biochem Mol Biol* 32, 205–210
25. Liu L, Laufer H, Wang Y, Hayes T (1997) A neurohormone regulating both methyl farnesoate synthesis and glucose metabolism in a crustacean. *Biochem Biophys Res Commun* 237, 694–701
26. Macins A, Meredith J, Zhao Y, Brock HW, Phillips IE (1999) Occurrence of ion transport peptide (ITP) and ion transport-like peptide (ITP-L) in Orthopteroids. *Arch Insect Biochem Physiol* 40, 107–118
27. Nassel DR (2002) Neuropeptides in the nervous system of *Drosophila* and other insects: multiple roles as neuromodulators and neurohormones. *Prog Neurobiol* 68, 1–84
28. Nielsen H, Soren B, von Heijen G (1999) Machine learning approaches for the prediction of signal peptides and other protein sorting signals. *Protein Eng* 12, 3–9
29. Riehle MA, Garczynski SF, Crim JW, Hill CA, Brown MR (2002) Neuropeptides and peptide hormones in *Anopheles gambiae*. *Science* 298, 172–175
30. Shultz JW, Regier JC (2000) Phylogenetic analysis of arthropods using two nuclear protein-encoding genes supports a crustacean + hexapod clade. *Proc R Soc Lond B* 267, 1011–1019
31. Soye D (1997) Occurrence and diversity of neuropeptides from the crustacean hyperglycemic hormone family in arthropods: a short review. *Ann NY Acad Sci* 814, 319–323
32. Spanings-Pierrot C, Soye D, van Herp R, Gompel M, Skaret G, Grousset E, Charmantier G (2000) Involvement of crustacean hyperglycemic hormone in the control of gill ion transport in the crab *Pachygrapsus marmoratus*. *Gen Comp Endocrinol* 119, 340–350
33. Tensen CP, Janssen KPC, van Herp R (1989) Isolation, characterization and physiological specificity of the crustacean hyperglycemic factors from the sinus gland of the lobster *Homarus americanus* (Milne-Edwards). *Invert Reprod Dev* 16, 155–164
34. Thompson JD, Gibson TJ, Plewniak R, Jeanmougin R, Higgins DG (1997) The Clustal X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 24, 4876–4882
35. Wainwright G, Webster SG, Wilkinson MC, Chung JS, Rees HH (1996) Structure and significance of mandibular organ-inhibiting hormone in the crab, *Cancer pagurus* involvement in multihormonal regulation of growth and reproduction. *J Biol Chem* 271, 12749–12754
36. Wang Y-J, Zhao Y, Meredith L, Phillips IE, Theilmann DA, Brock HW (2000) Mutational analysis of the C-terminus in ion transport peptide (ITP) expressed in *Drosophila* Kcl. cells *Arch Insect Biochem Physiol* 45, 129–138
37. Wheelan SJ, Chruch DM, Ostell JM (2001) Spidey: a tool for mRNA-to-genomic alignments. *Genome Res* 11, 1952–1957
38. Yasuda A, Yasuda Y, Fujita T, Naya Y (1994) Characterization of crustacean hyperglycemic hormone from the crayfish (*Procambarus clarkii*): multiplicity of molecular forms by stereoinversion and diverse functions. *Gen Comp Endocrinol* 95, 387–398